WHAT IS CLAIMED IS:

1. A reproducing method for receiving a stream of sent audio packets containing an audio code generated by encoding an input audio data stream frame by frame and reproducing an audio signal, comprising the steps of:

(a) storing received packets in a receiving buffer;

(b) detecting the largest delay jitter and the number of buffered packets, the largest jitter being any of the largest value and statistical value of jitter obtained by observing arrival jitter of the received packets over a given period of time and the number of buffered packets being the number of packets stored in the receiving buffer;

(c) obtaining, based on the largest delay jitter, an optimum number of buffered packets by using a predetermined relation between the largest delay jitter and the optimum number of buffered packets, the optimum number of buffered packets being the optimum number of packets to be stored in the receiving buffer;

(d) determining, on a scale of a plurality of levels, the difference between the detected number of buffered packets and the optimum number of buffered packets;

(e) retrieving a packet corresponding to the current frame from the receiving buffer and decoding an audio code in the packet to obtain a decoded audio data stream in the current frame; and

(f) performing any of expansion, reduction, and preservation of a waveform of the decoded audio data stream in accordance with a rule to make the number of buffered packets close to the optimum number of buffered packets, the rule being established for each level of the difference, and outputting the result as audio data of the current frame.

2. The reproducing method according to claim 1, wherein,

step (f) comprises the steps of:

(f-1) obtaining the pitch length of the decoded audio data stream;

(f-2) analyzing the audio data stream to determine whether the audio data stream is in a voice segment or a non-voice segment; and

(f-3) performing any of expansion, reduction, and preservation by inserting or removing a waveform corresponding to the pitch length in the decoded audio string or by not changing the decoded audio signal string, on the basis of the result of the determination of voice/non-voice segment and the result of the determination of the difference level.

3. The reproducing method according to claim 2, wherein,

step (d) comprises the step of determining whether the level of the difference represents a high urgency level indicating that the number of buffered packets should be urgently increased or decreased or a low urgency level indicating that the number of buffered packets should be slowly increased or decreased; and

step (f-3) comprises the step of, if the level represents the high urgency level, expanding or reducing the waveform of the decoded audio data stream regardless of whether the data stream is in a voice segment or a non-voice segment; if the level represents the low urgency level, expanding or reducing the waveform of the decoded audio data stream, on condition that the decoded audio data stream is in a non-voice segment.

4. The reproducing method according to claim 2, wherein,

step (d) comprises the step of determining whether the level of the difference represents a high urgency level indicating that the number of buffered packets should be urgently increased or decreased or a low urgency level indicating that the number of buffered packets should be slowly

increased or decreased;

step (f-3) comprises the step of, if the level represents the high urgency level, expanding or reducing the waveform of the decoded audio data stream regardless of whether the decoded audio data stream is in a voice segment or a non-voice segment, if the level represents the low urgency level, expanding or reducing the waveform of the decoded audio data stream once every predetermined number N1 of frames on the condition that the decoded audio data stream is in a voice segment, or expanding or reducing the waveform of the decoded audio data stream once every predetermined number N2 of frames on the condition that the decoded audio data stream is in a non-voice period, where N1 and N2 being integers greater than or equal to 1 and N2 is smaller than N1.

5. The reproducing method according to claim 1, wherein,

step (f) comprises the steps of:

(f-1) obtaining the pitch length of the decoded audio data stream;

(f-2) analyzing the decoded audio data stream to determine which of a voiced sound segment, an unvoiced sound segment, a background noise segment, and a silence segment the decoded audio data stream is in;

(f-3) performing any of expansion, reduction, and preservation of the decoded audio data stream by inserting or removing a waveform corresponding to the pitch length in the decoded audio data stream or by not changing the decoded audio data stream, on the basis of the result of the segment determination and the result of the determination of the difference level.

6. The reproducing method according to claim 5, wherein,

step (d) comprises the step of determining whether the level of the difference represents a high urgency level indicating that the number of

buffered packets should be urgently increased or decreased or a low urgency level indicating that the number of buffered packets should be slowly increased or decreased; and

step (f-3) comprises the step of, if the level represents the high urgency level, expanding or reducing the waveform of the decoded audio data stream regardless of the result of the segment determination; if the level represents a low urgency level, expanding or reducing the waveform of the decoded audio data stream once every predetermined number N1, N2, N3, N4 of frames, the predetermined number being predetermined for each of the voiced sound segment, the unvoiced sound segment, the background noise segment, and the silence segment, where N1, N2, N3, and N4 are positive integers and at least one of the integers is greater than or equal to 2 and differs from the other three integers.

7. A reproducing apparatus for audio packets which receives a stream of sent audio packets containing an audio code generated by encoding an input audio data stream frame by frame and reproduces an audio signal, comprising:

a packet receiving part which receives audio packets from a packet communication network;

a receiving buffer for temporarily storing the received packets and reading out packets in response to a request;

a state detecting part which detects the largest delay jitter and the number of buffered packets, the largest jitter being any of the largest value and statistical value of jitter obtained by observing arrival jitter of the received packets over a given period of time and the number of buffered packets being the number of packets stored in the receiving buffer;

a control part which obtains based on the largest delay jitter an

optimum number of buffered packets by using a predetermined relation between the largest delay jitter and the optimum number of buffered packets, the optimum number of buffered packets being the optimum number of packets to be stored in the receiving buffer, determines, on a scale of a plurality of levels, the difference between the detected number of buffered packets and the optimum number of buffered packets, and generates a control signal for instructing to perform any of expansion, reduction, and preservation of a waveform of the decoded audio data stream in accordance with a rule to make the number of buffered packets close to the optimum number of buffered packets, the rule being established for each level of the difference;

an audio packet decoding part which decodes an audio code in a packet corresponding to the current frame extracted from the receiving buffer to obtain a decoded audio data stream in the current frame; and

a consumption adjusting part which performs any of expansion, reduction, and preservation of the waveform of the decoded audio data stream in accordance with the control signal and outputs the result as sound data of the current frame.

8. The reproducing apparatus according to claim 7, further comprising an audio analyzing part analyzes the decoded audio data stream to determine whether the decoded audio data stream is in a voice segment or a non-voice segment, provides the result of the determination to the control part, obtains the pitch length of the decoded audio data stream, and provides the pitch length to the consumption adjusting part; wherein,

the control part provides control to cause the consumption adjusting part to perform any of expansion, reduction, and preservation of the decoded audio data stream of the current frame, on the basis of the result of the segment determination and the result of the difference level determination;

and

the consumption adjusting part inserts or removes a waveform corresponding to the pitch length in the decoded audio data stream or does not change the decoded audio data stream, in accordance with the control.

9. The reproducing apparatus according to claim 8, wherein the control part determines whether the level of the difference represents a high urgency level indicating that the number of buffered packets should be urgently increased or decreased or a low urgency level indicating that the number of buffered packets should be slowly increased or decreased; and, if the level represents the high urgency level provides control to cause the consumption adjusting part to expand or reduce the waveform of the decoded audio data stream regardless of whether the data stream is in a voice segment or a non-voice segment; if the level represents the low urgency level, provides control to cause the consumption adjusting part to expand or reduce the waveform of the decoded audio data stream, only on condition that the decoded audio data stream is in a non-voice segment.

10. The reproducing apparatus according to claim 8, wherein the control part determines whether the level of the difference represents a high urgency level indicating that the number of buffered packets should be urgently increased or decreased or a low urgency level indicating that the number of buffered packets should be slowly increased or decreased; and, if the level represents the high urgency level, provides a control to cause the consumption adjusting part to expand or reduce the waveform of the decoded audio data stream regardless of whether the decoded audio data stream is in a voice segment or a non-voice segment; if the level represents the low urgency level, provides a control to cause the consumption adjusting part to expand or reduce the waveform of the decoded audio data stream once every

predetermined number N1 of frames on the condition that the decoded audio data stream is in a voice segment, or to expand or reduce the waveform of the decoded audio data stream once every predetermined number N2 of frames on the condition that the decoded audio data stream is in a non-voice period, where N1 and N2 being integers greater than or equal to 1 and N2 is smaller than N1.

11. The reproducing apparatus according to claim 7, wherein the audio analyzing part analyzes the decoded audio data stream to determine which of a voiced sound segment, an unvoiced sound segment, a background noise segment, and a silence segment the decoded audio data stream is in, provides the result of the determination to the control part, obtains the pitch length of the decoded audio data stream, and provides the pitch length to the consumption adjusting part;

the control part provides a control based on the result of the segment determination and the result of the difference level determination to the consumption adjusting part to perform any of expansion, reduction, and preservation of the decoded audio data stream of the current frame; and

the consumption adjusting part, in accordance with the control, inserts or removes a waveform corresponding to the pitch length in the decoded audio data stream or does not change the decoded audio data stream.

12. The reproducing apparatus according to claim 11, wherein the control part determines whether the level of the difference represents a high urgency level indicating that the number of buffered packets should be urgently increased or decreased or a low urgency level indicating that the number of buffered packets should be slowly increased or decreased; and, if the level represents the high urgency level, provides a control to cause the consumption adjusting part to expand or reduce the waveform of the decoded

audio data stream regardless of the result of the segment determination; if the level represents a low urgency level, provides a control to cause the consumption adjusting part to expand or reduce the waveform of the decoded audio data stream once every predetermined number N1, N2, N3, N4 of frames, the predetermined number being predetermined for each of the voiced sound segment, the unvoiced sound segment, the background noise segment, and the silence segment, where N1, N2, N3, and N4 are positive integers and at least one of the integers is greater than or equal to 2 and differs from the other three integers.

13. A reproducing program for audio packets written in a computer-interpretable language for causing a computer to perform the reproducing method according to claim 1.

14. A recording medium formed by a computer-readable recording medium and having recorded thereon the reproducing program according to claim 13.